DOCUMENT RESUME

ED 156 640 SP 012 748

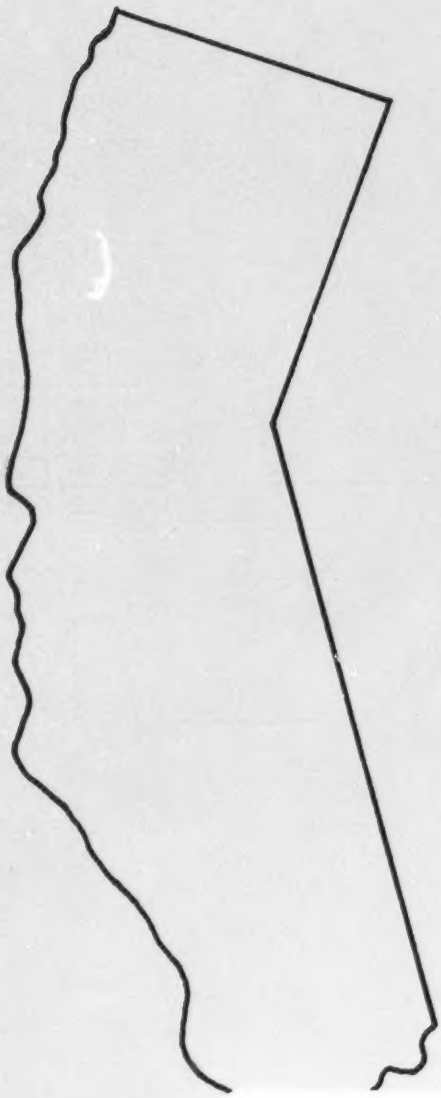
TITLE	Design for Use of Individual Testing in the Study of Instructional Time. Beginning Teacher Evaluation Study, Technical Note Series. Technical Note III-2.
INSTITUTION	Far West Lab. for Educational Research and Development, San Francisco, Calif.
PUB DATE	15 Sep 75
NOTE	17p.; For related documents, see SP 012 741-751
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Educational Research; Elementary Education; Grade 2; Grade 5; *Group Norms; *Group Tests; Mathematics; Performance Factors; Program Planning; Reading; Test Reliability
IDENTIFIERS	*Test Format

ABSTRACT

This report describes the general plan for a study to clarify the interpretation of group performance measures of student achievement. Group tests are the most efficient procedure for collecting achievement information from a large number of students, but they also require item formats that may be different from the form in which students normally demonstrate or apply their learning. To provide information about some of these potential differences to be used in the interpretation of group scores, two studies are planned, one of second grade students in the areas of reading and mathematics, and one of second and fifth grade students in reading comprehension. (MJB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED156640



BTES

BEGINNING TEACHER EVALUATION STUDY

Technical Note Series

TECHNICAL NOTE III-2

Design for Use of Individual Testing in the Study of Instructional Time

September 15, 1975

SP012748

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
1855 FOLSOM STREET • SAN FRANCISCO, CALIFORNIA 94103 • (415) 565-3000

TECHNICAL NOTE III-2

CONTINUATION OF PHASE IIIA

Design for Use of Individual Testing in the Study of Instructional Time

September 15, 1975

CONTENTS

	PAGE
Introduction	1
Study 1	3
Subjects	3
Procedure	3
Math Tasks*	4
Reading Tasks*	6
Study 2	8
Subjects	8
Procedure	9
Reading Tasks*	9
Reports Scheduled	10
References	12

*Overview. For details see Technical Note III-3.

Introduction

Task III of Phase IIIA Extension of the Beginning Teacher Evaluation Study is the refinement of a student test battery in reading and mathematics for grades 2 and 5. The major part of the refinement effort is directed toward a set of group-administered paper-and pencil tests. A set of tests was developed during Phase IIIA and is being pilot tested during Phase IIIA Extension. These group tests, along with a discussion of procedures for development and refinement, are presented in Technical Report III-1 (Filby and Dishaw, 1975).

An additional component of the Task III test development effort is the development of individually administered tests. The major purpose of the individual test effort is to validate the interpretation of group-test measures. We want to be sure that we understand what the group-test measures are measuring. Group-administered paper-and-pencil tests must be used in a large scale study as the major means of assessing student performance levels. But the format of such tests places a number of limitations on the kind of test items which can be used. Often the most direct means of assessing a student's skill become impossible to use under group-test conditions. For example, to assess a student's decoding ability, one might ask the student to read words aloud. The student's specific deficiencies would quickly become apparent in such a task. In a group test, however, the child cannot give an oral response. As a consequence, alternative testing formats are being developed which are intended to tap the same underlying skill by more direct measures. Performance on group measures may be influenced by a number of factors in addition to the skill of primary interest. Multiple interpretations of scores on group-test measures are usually possible and are naturally undesirable.

In order to clarify the interpretation of group performance measures, a subset of the children involved in the group-test pilot study will also be tested individually on selected tasks. These tasks will be designed to parallel the content of subtests in the group-test battery. A comparison can then be made between performance on the group-test measures and performance on related individually administered tasks. If there is substantial agreement between the two measures, we can be more confident in our interpretation of group-test scores. If there is noticeable disparity between allegedly parallel measures, then we will need to reconsider the group-test measures. It may be that selective individual testing measures should be used to provide important information to supplement group-test assessment.

Individual testing will be carried out in two separate studies. Study 1 will involve grade two students only, and will include both reading and mathematics exercises. Study 2 will focus specifically on reading comprehension measures and will be carried out with students from both grade 2 and grade 5.

The design and general procedure for each study will be described in Technical Note (III-2). More specific information about materials, instructions, and scoring are found in Technical Note III-3.

Study 1

Subjects. Of the 15 grade 2 classes involved in the pilot testing of the group tests, 6 classes are being observed extensively as part of a study of instructional time (Task II; see Technical Note II-1). All 6 of these classes along with 2 additional non-observed classes, will receive the set of reading subtests which emphasize decoding. From these 8 classes, 30 children will be selected for individual testing. This overlap with the instructional time study will mean that analyses of instructional time effects can be carried out using test data from individual testing as the dependent measures. This analysis with a subset of the data would allow a comparison of the degree of success in predicting achievement using the two types of measures.

Procedure. Each of the 30 students will be tested in October and December. Each individual testing period will take place within two weeks of the corresponding group test session. Each individual testing session will take from 30 to 40 minutes, with half the time spent on reading and half on mathematics. If the student becomes tired or restless, a break will be taken between the two parts. The concern throughout the test sessions is to provide and maintain a relaxed and non-threatening atmosphere. We want the child to feel encouraged to respond freely.

All testing will be carried out by trained staff members familiar with the group test development. Pilot testing of materials and procedures will be carried out for several weeks prior to data collection. Training sessions will be held so that testers can discuss interview techniques and standardize procedures. This pilot testing will provide an important opportunity to refine and tighten the individual measures. Care will be taken to identify

decision rules for routing children through the testing sequence.

All sessions will be audio tape recorded. In addition, testers will keep notes about student performance. All notes, materials, and tapes will be returned to the Laboratory for coding and data analysis.

The set of tasks to be covered in each testing session is shown in Figure 1. An overview and rationale for task selection are given in the following sections.

Mathematics Tasks. The understanding of place value is one of the most important objectives in second grade mathematics. To understand the operations of our number system, a child must first fully understand the structure of our numeration system in that it is a decimal system in which the 10 plays the key role. The child needs to understand that the face value and relative position of a digit in a numeral determine its total value. The understanding of place value is essential to the development of regrouping skills in addition and subtraction. Because of its centrality in second grade mathematics, place value has been selected as the focal point for the individual testing.

The grade 2 test battery for Phase III-A contains a 15 item place value scale. A comparison of individual and group test results will give us an indication of the appropriateness of the group test format in testing the child's understanding of place value. Due to the limitations of paper and pencil tests the items in the place value scale can only be presented on semi-concrete (pictorial representation) and abstract levels. It is generally believed that conceptual understanding is best facilitated initially by the use of concrete materials. At the time of the individual testing most children will still be in the process of acquiring the concept of place value and therefore the testing will employ manipulative materials in an effort to tap the child's current level of conceptual understanding. The use of

concrete materials will be combined with problems involving the abstract representation of numbers in a graduated sequence of tasks.

Figure 1

Study 1 Tasks

	October Test (30-40 minutes)	December Test (30-40 minutes)
Math	1. Ordering Numerals	1. Ordering Numerals
	2. Renaming 10	2. Renaming 10 (reduced scale)
	3. Place Value	3. Place Value
	4. Addition with Regrouping	4. Addition with Regrouping
		5. Subtraction with Regrouping
Reading		1. Sight Words
		2. Word List
		3. Words in Context
	4. Long-Short Vowels--Nonsense Words	4. Long-Short Vowels--Nonsense Words
	5. Passages--Oral Reading and Comprehension	5. Passages--Oral Reading and Comprehension

The child will use the manipulatives to solve a problem, to explain his thought processes, and to show or prove his written solution to a problem. At this age level it is very difficult for a child to verbalize a mathematical process or operation at the abstract level but verbalization should be easier when directly associated with concrete materials. The child's use of the materials and his verbal explanations should demonstrate whether he understands place value and whether this understanding is on a rational or mechanical level.

Five component tasks are being investigated in the individual testing. They can be labeled as follows:

(1) Ordering the Numerals 1 through 10
(2) Renaming the Numeral 10
(3) Place Value
(4) Addition with Regrouping
(5) Subtraction with Regrouping

Reading Tasks. The individual tasks in reading for grade 2 focus on the validation of the decoding scales in the group test battery. As discussed in the introduction, the most direct measure of a child's word attack skills is to ask him to read a set of words aloud. Oral reading of words, sentences, and stories, is the major activity in the individual testing. Error patterns on several oral reading tasks can then be compared to performance on group test measures on decoding subskills. The following comparisons are of particular interest.

(1) The primary format of the decoding subtests is to present a picture and ask the reader to select the one of three words which correctly names the picture. The response choices are identical except for a specific feature. A sample group test item is the following:

The focus of attention in this item is on long and short vowel sounds. The hypothesis is that students who do well on this picture naming task will also be able to read words containing a long vowel sound in the silent e pattern. This hypothesis will be tested by asking the student to read a set of words like kite; words which are correct responses in group test items. Ability to generalize phonics knowledge will also be tested by asking the student to read "nonsense" words. Particular attention is paid to the long vowel patterns, because they are the focus of content coding during observation in the instructional time study (Task II). Other group test items will also be represented in the word list for oral reading.

(2) The Sight Words subtest of the group test battery uses another common format. The tester pronounces a word and the student must find that word in a set of five choices. For example, the word where must be identified within the set "where, whate, were, whose," Again the hypothesis is that the ability to identify the correct word is highly correlated with the ability to read the word aloud. Again the hypothesis will be tested by asking each student to read aloud the list of target sight words.

(3) Both of the comparisons above deal with words in isolation. In the normal process of reading, words are embedded in a context and understanding of that context aids in word identification. Particularly for common function words such as those in the Sight Words test, the skills needed to read the word in isolation may be very different from those needed to read the word in context. Because of the essential importance of reading normal prose, both group test and individual measures of word identification will be compared to error patterns in oral reading of prose. If word identification tasks legitimately diagnose decoding deficiencies, then the same types of errors should show up when students read prose passages.

(4) A final comparison looks specifically at the ability to process and use sentence context to guide decoding. Several subtests in the group battery examine the ability to use context clues. In individual testing, sentence contexts will be provided for all words in the word list for oral reading. If a student cannot read the word in isolation, he will be given the context. The number of additional words read in context is a good measure of the ability to use context to identify unknown words. A second measure of context processing will be derived from an analysis of oral reading of passages. Even when a reader makes an error, he should tend to produce utterances which are grammatical and meaningful. If anomalous statements are produced and not corrected, this is an indication that the reader is not processing language units. Again, the interpretation of the group test measures of context clues will be greatly enhanced by a comparison with sentence context and passage reading.

To summarize tasks, the following five components will be included in individual testing of reading for grade 2:

(1) Sight Words
(2) Word List
(3) Words in Context
(4) Long-Short Vowels--Nonsense Words
(5) Passages--Oral Reading and Comprehension

Study 2

Subjects. A separate sample of students will be used for Study 2, which examines the reading comprehension measures in the group test. Several classes not chosen for participation in the pilot testing of the group test battery will be invited to participate in this smaller study. Two classes at each grade level will be recruited, with a minimum of 30 students tested per grade. The use of a separate sample is considered necessary because the intensive interview might influence subsequent performance on the group test and thereby interfere with the pilot test study. The sample chosen for this

special study should be comparable to the pilot study sample.

Procedure. Early in January the reading comprehension passages will be administered in a group setting to participating classes. One 45-minute session should be sufficient for this purpose. Within the following two weeks, students will be individually tested on the same passages. This individual session will function like a debriefing. Students will read the passage orally, then read and respond to each comprehension question. Students will be asked to explain how the answer was selected and why other choices are incorrect. Additional probe questions may be used to gather further information about the reader's understanding of the passage.

As with Study 1, training sessions will be held to standardize the interview procedure. Audio tape recording of sessions will be used in order to do thorough and accurate coding of responses.

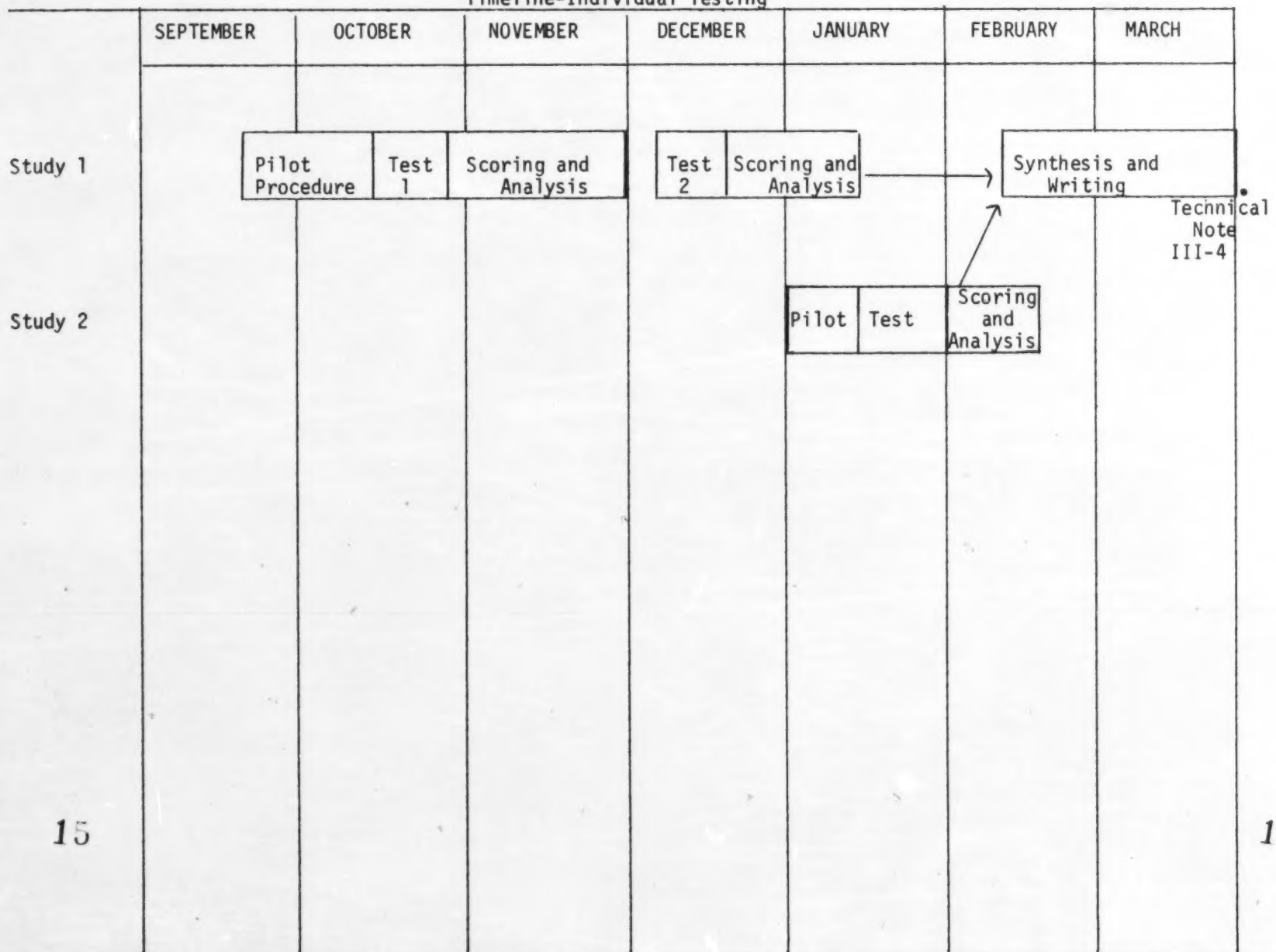
Reading Tasks. A single reading task--probe comprehension--provides the focus for this study, and the task is largely explained in the procedure section above. Passages and questions will be the same as those used in the group test. Special attention will be paid to items which require inference or depend on background knowledge. Questions and/or answer choices which are ambiguous and misunderstood will be revised or eliminated. Choices which are too plausible, which mislead imaginative readers, will also be reconsidered.

Reports Scheduled

The timelines for both studies are shown in Figure 2. The two strands merge in the spring, when results of both studies are presented in Technical Note III-4. This Note will examine the results of the individual testing and the implications for the interpretation of the group test measures. Recommendations about the revision and/or extension of the group test battery will be made. It is believed that the assessment of student achievement in Phase IIIB will be strengthened by this investigation of the meaning of the student achievement measures.

Figure 2

Timeline-Individual Testing



References

1. Design for a study of instructional time. Technical Note II-1, BTES Technical Report Series, Far West Laboratory, August, 1975.

2. Filby, N. and Dishaw, M. Development and refinement of reading and mathematics tests for grades 2 and 5. Technical Report III-1, BTES Technical Report Series, Far West Laboratory, August, 1975.

3. Instruments for individual Assessment of Achievement. Technical Note III-3, BTES Technical Report Series, Far West Laboratory, September, 1975.